

Detecting compact galactic binaries using a hybrid swarm-based algorithm

Yann Bouffanais^{1,*} and Edward K. Porter^{1,†}

¹*François Arago Center, APC, UMR 7164, Université Paris 7 Denis Diderot,
10, rue Alice Domon et Léonie Duquet, 75205 Paris Cedex 13, France*

Compact binaries in our galaxy are expected to be one of the main sources of gravitational waves for the future eLISA mission. During the mission lifetime, many thousands of galactic binaries should be individually resolved. However, the identification of the sources, and the extraction of the signal parameters in a noisy environment are real challenges for data analysis. So far, stochastic searches have proven to be the most successful for this problem. In this work we present the first application of a swarm-based algorithm combining Particle Swarm Optimization and Differential Evolution. These algorithms have been shown to converge faster to global solutions on complicated likelihood surfaces than other stochastic methods. We first demonstrate the effectiveness of the algorithm for the case of a single binary in a 1 mHz search bandwidth. This interesting problem gave the algorithm plenty of opportunity to fail, as it can be easier to find a strong noise peak rather than the signal itself. After a successful detection of a fictitious low-frequency source, as well as the verification binary RXJ0806.3+1527, we then applied the algorithm to the detection of multiple binaries, over different search bandwidths, in the cases of low and mild source confusion. In all cases, we show that we can successfully identify the sources, and recover the true parameters within a 99% credible interval.

PACS numbers:

I. INTRODUCTION

Compact binary systems, composed of white dwarves (WDs), neutron stars (NSs) and stellar mass black holes (BHs), are thought to be a major source of gravitational waves (GWs) in our galaxy. Estimations from population synthesis models predict a possible number of close to 70 million galactic binaries in the data at any one time [1, 2]. The binaries of interest for space-based GW astronomy are composed of two compact stars with orbital periods shorter than one hour. These binaries can be sorted into two categories: detached and interacting. In the case of detached binaries, there is no mass transfer between the two objects, and their dynamics is essentially governed by GW emission. Most galactic binary (GB) systems will be composed of binary WDs, but there will also be binary NS and binary BH systems. We should also expect a number of mixed binaries composed of different permutations compact object type. Interacting binaries are characterized by a transfer of momentum via mass accretion. GWs emitted by these objects would give crucial information on their complex dynamics. There are two main categories of interacting binaries: AM CVn systems where the accretor and the companion are WDs, and ultra compact x-ray binaries where the accretor is a NS, while the donor is a WD [3]. Due to their weak interaction with matter, the observation of these objects with GWs should give us new insights into their nature and would add complementary information to the current electromagnetic (EM) observations. In fact, the GWs

emitted by these sources carry important information on the distances, masses, and nature of the binary system (i.e. detached or interacting with a change in frequency). All of these observations can then be used in order to better understand the formation and distribution of compact objects in our Galaxy [4, 5].

However, identifying the parameters of all of these simultaneous sources is a real challenge in terms of data analysis. The first difficulty is the weak interaction between GWs and the detector, leading to very small wave amplitudes. As a consequence, all the signals will be hidden in the noise of the instrument, and only a Fourier analysis reveals the features of the signal in the noise background. Another issue is closely connected to the nature of gravitational waves and is often called the *Cocktail party problem* [6]. Unlike EM observations, a GW detector measures the signals in all directions simultaneously. The advantage of this simultaneous detection is the ability to constantly observe and gather information on the sources. The drawback is the confusion between signals: two binaries with extremely close frequencies cannot be analyzed separately if the difference between their frequencies is smaller than the smallest accessible frequency bin $\Delta f = 1/T_{obs}$, where T_{obs} is the length of the observation period. Thus, we know that in certain frequency bins, there will be more than one binary, and it will be impossible to separate the two signals: only the one with the highest SNR will be recovered. While source confusion is a problem for the eLISA mission, we do not expect to have the same confusion noise from a GB foreground as with the LISA configuration due to the reduced size of the observatory [7]. Furthermore, our ability to individually resolve sources comes from the fact that the power in a single binary is spread over several frequency bins due to the Doppler modulation coming from the orbit of the

*bouffana@apc.in2p3.fr

†porter@apc.in2p3.fr

detector around the Sun [8]. This modulation frequency is denoted as $f_m = 1/yr$.

Regarding these challenges, it has been necessary to develop algorithms dedicated specifically to parameter identification for GWs. There have been two major types of algorithms tested in the framework of space-based GW data analysis: grid based and stochastic searches. Grid-based algorithms use a uniform distribution of templates over the parameter space, designed to achieve a given degree of minimal match with the sources. They can either be in the form of a hierarchical search, where regions of interest are identified separately with refinement of the grid [9], or in the form of an N-source search where metatemplates of multiple binaries are computed [10]. However, given the high computational cost of computing a uniform grid of templates in a high-dimension parameter space [11], stochastic approaches have been favored in the field. A variety of such algorithms have been tested, including genetic algorithms [12], tomographic reconstruction [13] and Markov chain Monte Carlo based algorithms [14–22]. Among these algorithms, the MCMC-based pipelines have been the most successful so far, especially the Block Annealed Metropolis Hastings that proved to be the most suited for detecting a full population of binaries in the Galaxy [15].

Though much progress has been made in the GB source search, it is still necessary to look for even faster and more efficient algorithms in order to be ready to analyse real data when the mission is launched. In this article we present a first work on a swarm-based search algorithm that combines Particle Swarm Optimization (PSO) [23, 24] and Differential Evolution (DE) [25] during the search phase, as well as DE and Markov Chain Monte Carlo (MCMC) [26, 27] routines for the parameter estimation phase. While MCMC-based algorithms have performed well, their convergence is very much dependent on the choice of an efficient proposal distribution and annealing strategies. Once we are close to the global solution, the convergence of these methods is quite fast. However, at a distant point in parameter space, these algorithms may find themselves stuck in a random walk with slow convergence. Both PSO and DE have been shown to have faster convergence properties than MCMC methods [28, 29], especially when the likelihood surface is known to have a high population of secondary maxima.

eLISA is a future ESA L3 project dedicated to the detection of GWs using space-based interferometry [7]. So far there has been no direct detection of GWs, but only indirect evidence from observations of binary pulsars [30]. The eLISA mission is being constructed as a complementary observatory to current Earth-based projects such as Advanced LIGO and Advanced VIRGO [31, 32], as well as pulsar timing arrays [33]. The eLISA frequency bandwidth of detection is set between 10^{-5} and 1 Hz. In this bandwidth, we should detect a number of sources such as GBs, supermassive black hole binaries, extreme mass ratio inspirals, cosmic (super)strings, and possibly, a stochastic cosmological background. Among these

sources, the vast majority will be GBs. In its current configuration, eLISA is a three-body constellation defining a two-arm interferometer of length $L = 10^6$ km. The constellation will orbit the Sun at a distance of 1 AU and with an angle of 20° with respect to the Earth. Due to the cartwheel movement of the spacecrafts, various modulations play a part in the phase of the signals measured by the detector. In terms of GB sources, their frequencies span the band between 10^{-4} and 10^{-2} Hz, and it is believed that the number of resolvable galactic binaries will be close to 4000 in the first year, with an additional 10^3 per year of observation [7, 34].

This article is structured as follows. In Sec. II, we review the important concepts of Bayesian inference and their application to the detection of gravitational waves for ultra-compact galactic binaries. In Sec. III, we present the general features of the methods that have been used in our search algorithm. In Sec. IV, we present the design of our search algorithm and how we benchmarked its performance on a single source search. In Sec. V, we present the performances of our search algorithm for the analysis of two multiple sources data sets.

II. DATA ANALYSIS METHODOLOGY FOR GRAVITATIONAL WAVE ASTRONOMY

The main difficulty in GW astronomy comes from the fact that the output of the detector is a superposition of millions of GW signals, plus a number of noise contributions. A standard technique for extracting signals from noisy data is matched filtering. The idea behind this method is as follows : given a theoretical waveform model, or template, parametrized by astrophysically motivated parameters $\{\lambda^\mu\}$, what is the parameter set that maximizes the correlation between the template and a possible individual GW signal in the data?

Once a signal is detected, our next goal is to estimate the parameters for the best fit template. A number of recent studies have demonstrated that the Fisher information matrix is an unreliable tool in GW astronomy due to the fact that in a number of cases, the posterior distributions in the parameter errors are non-Gaussian (a prerequisite for using the Fisher matrix in the first place) [35]. With this in mind, our goal is to carry out a Bayesian analysis when conducting the parameter estimation study.

A. A Bayesian framework for data analysis

If we consider a time-series signal $s(t)$ coming from the output of an experiment, we assume that $s(t)$ contains both a noise-like part $n(t)$, related to noise coming from experimental or physical backgrounds, and a signal of interest $h(t; \lambda^\mu)$. This signal depends on a parameter set $\{\lambda^\mu\}$ that describes the physics of the phenomenon.

Thus, we can write

$$s(t) = h(t; \lambda^\mu) + n(t). \quad (1)$$

Bayesian inference is quickly becoming a useful tool in GW astronomy for testing model hypothesis, i.e. given a data set $s(t)$, a template $h(t; \lambda^\mu)$, and a noise model $S_n(f)$, what is the posterior probability distribution $p(\lambda^\mu | s)$? One can construct the posterior probability distribution via Bayes' theorem:

$$p(\lambda^\mu | s) = \frac{\pi(\lambda^\mu) p(s | \lambda^\mu)}{p(s)}. \quad (2)$$

The prior probability $\pi(\lambda^\mu)$ reflects the *a priori* knowledge of the experiment before evaluating the data, $p(s | \lambda^\mu) = \mathcal{L}(\lambda^\mu)$ is the likelihood function, and $p(s)$ is the marginalized likelihood or “evidence” given by

$$p(s) = \int d\lambda^\mu \pi(\lambda^\mu) p(s | \lambda^\mu). \quad (3)$$

We define the likelihood, $\mathcal{L}(\lambda^\mu)$, given the output, $s(t)$, and a GW template, $h(t)$, as

$$\mathcal{L}(\lambda^\mu) = \exp \left[-\frac{1}{2} \langle s - h(\lambda^\mu) | s - h(\lambda^\mu) \rangle \right], \quad (4)$$

where the angular brackets define a noise-weighted scalar product defined on the manifold of all possible signals. More explicitly, given two time domain signals $h(t; \lambda_1^\mu)$ and $s(t; \lambda_2^\mu)$, we can write the scalar product as

$$\langle h | s \rangle = 2 \int_0^\infty \frac{df}{S_n(f)} \left[\tilde{h}(f) \tilde{s}^*(f) + \tilde{h}^*(f) \tilde{s}(f) \right], \quad (5)$$

where $\tilde{h}(f)$ is the Fourier transform of the time domain signal $h(t)$, an asterisk denotes a complex conjugate and $S_n(f)$ is the one-sided noise power spectral density of the noise. In this study, we use the eLISA observatory as our detector of choice. In its current configuration, the one-sided noise power spectral density is given by [7]

$$\begin{aligned} S_n(f) = & \frac{1}{4L^2} \left[S_n^{fxd} + 2S_n^{\text{pos}} \left(2 + \cos^2 \left(\frac{f}{f_*} \right) \right) \right. \\ & + 8S_n^{\text{acc}} \left(1 + \cos^2 \left(\frac{f}{f_*} \right) \right) \\ & \left. \times \left(\frac{1}{(2\pi f)^4} + \frac{(2\pi 10^{-4})^2}{(2\pi f)^6} \right) \right], \end{aligned} \quad (6)$$

where $L = 10^9 \text{m}$ is the arm length of the observatory, $S_n^{fxd} = 6.28 \times 10^{-23} \text{m}^2/\text{Hz}$ is a frequency independent fixed level noise in the detector, $S_n^{\text{pos}}(f) = 1.21 \times 10^{-22} \text{m}^2/\text{Hz}$ is the position noise and $S_n^{\text{acc}}(f) = 9 \times 10^{-30} \text{m}^2/(\text{s}^4 \text{Hz})$ is the acceleration noise. We highlight the fact that our expression for the noise also contains a red-noise component at frequencies of $f \leq 10^{-4} \text{Hz}$.

Complementary to the likelihood function, we can also define the signal-to-noise ratio (SNR) as

$$\rho = \frac{\langle s | h \rangle}{\sqrt{\langle h | h \rangle}}. \quad (7)$$

B. Time domain response for galactic binaries

The strain of the GW as seen by the detector can be written, in the low frequency approximation (LFA) [36], as a linear combination of the two GW wave polarizations $h_{+, \times}(t)$ and the detector beam pattern functions $F^{+, \times}(t)$,

$$h(t) = h_+(t)F^+(t) + h_\times(t)F^\times(t). \quad (8)$$

The LFA is valid when the GW wavelength is greater than the size of the detector, or conversely, where the frequency of the wave is inferior to the mean transfer frequency $f^* = c/(2\pi L)$ [37, 38]. For eLISA, this corresponds to a frequency of $f^* \approx 10^{-2} \text{Hz}$.

In the case of circular monochromatic GBs, and in the framework of general relativity, the two polarizations of the GW in Eq. (8) are given by

$$h_+(t) = A(1 + \cos^2(\iota)) \cos(\Phi(t) + \varphi_0), \quad (9)$$

$$h_\times(t) = -2A \cos(\iota) \sin(\Phi(t) + \varphi_0), \quad (10)$$

where A is the amplitude of the wave, ι is the inclination of the orbital plane, φ_0 is the initial phase and $\Phi(t)$ is the phase of the wave. For a monochromatic binary, the phase can be expressed as

$$\Phi(t) = 2\pi f_0 [t + R_\oplus \sin(\theta) \cos(2\pi f_m t - \phi)], \quad (11)$$

where f_0 is the GW frequency (twice the value of the orbital frequency), θ is the colatitude, ϕ is the longitude and $R_\oplus = 1 \text{AU}$ is the radius of eLISA orbit. The phase differs from a simple monochromatic process due to the motion of the detector with respect to the source, inducing a Doppler motion contribution to the phase. The beam pattern functions of the detector, $F^{+, \times}(t)$, are described in the LFA by

$$\begin{aligned} F^+(t; \psi, \theta, \phi) = & \frac{1}{2} [\cos(2\psi) D^+(t; \psi, \theta, \phi, \lambda) \\ & - \sin(2\psi) D^\times(t; \psi, \theta, \phi, \lambda)], \end{aligned} \quad (12)$$

$$\begin{aligned} F^\times(t; \psi, \theta, \phi) = & \frac{1}{2} [\sin(2\psi) D^+(t; \psi, \theta, \phi, \lambda) \\ & + \cos(2\psi) D^\times(t; \psi, \theta, \phi, \lambda)], \end{aligned} \quad (13)$$

where ψ is the polarization angle of the wave and the full expressions for the coefficients $D^{+, \times}(t)$ are given in [37]. In general, for a circular monochromatic binary, we characterize the GW response by the set of seven parameters λ^μ :

$$\lambda^\mu = \{\ln A, \cos i, \varphi_0, \psi, \ln f_0, \cos \theta, \phi\}. \quad (14)$$

C. Fourier domain response for galactic binaries

It is also possible to accelerate the waveform generation by working directly in the Fourier domain. As

well as having an acceleration in the waveform generation from not needing to generate long arrays for the time domain waveforms, nor having to evaluate numerical Fast Fourier Transforms, a further acceleration is gained due to the fact that, for galactic compact binaries the waveform power is packed inside a small number of frequency bins. This procedure has already been derived and tested in other works in the framework of compact galactic binaries [8, 21, 39], so here, we present only the main concepts behind the formulation.

We can rewrite the expression for the time domain response as

$$s(t) = A_+ F^+ \cos(\Phi(t)) + A_\times F^\times \sin(\Phi(t)), \quad (15)$$

where $A_+ = A(1 + \cos^2(\iota))$ and $A_\times = -2A \cos(\iota)$. The phase Φ can be further be written as the sum of three contributions

$$\Phi(t) = 2\pi f_0 t + \Phi_D(t) + \varphi_0, \quad (16)$$

where we have a binary frequency term, a Doppler modulation term, and the initial phase. To find the analytic expressions for the Fourier coefficients, it is easier to express the response in the form

$$s(t) = \text{Re} \left[A_+ F^+ e^{2\pi i f_0 t} e^{i\Phi_D(t)} e^{i\varphi_0} \right] + \text{Im} \left[A_\times F^\times e^{2\pi i f_0 t} e^{i\Phi_D(t)} e^{i\varphi_0} \right], \quad (17)$$

where each term inside the brackets needs to be represented in the Fourier domain for an arbitrary observation time T_{obs} .

Dealing with each term in sequence, first of all, given the LFA, the detector functions can be easily expressed as a finite Fourier series of harmonic functions of the modulation frequency f_m

$$F^{+, \times}(t) = \sum_{k=-4}^4 p_k^{+, \times} e^{2\pi i f_m k t}, \quad (18)$$

where the harmonic coefficients $p_k^{+, \times}$ can be deduced from the expressions for $D^{+, \times}$ [37]. Now since the observation time T_{obs} is arbitrary, the associated Fourier coefficients of the detector functions $\tilde{p}_n^{+, \times}$ are obtained via the convolution of the above expression with the Fourier transform of a step function, which takes values between 0 and T_{obs} ; i.e.,

$$\tilde{p}_n^{+, \times} = \sum_{k=-4}^4 p_k^{+, \times} \text{sinc}(\pi(k f_m T_{obs} - n)) e^{i\pi(k f_m T_{obs} - n)}, \quad (19)$$

where we define the cardinal sine function $\text{sinc}(x) = \sin(x)/x$. In the same way, one can find the Fourier coefficients \tilde{a}_n associated with the monochromatic binary frequency term of the phase by a convolution as

$$\tilde{a}_n = \text{sinc}(\pi(f_0 T_{obs} - n)) e^{i\pi(f_0 T_{obs} - n)}, \quad (20)$$

where the power contribution is peaked around the carrier frequency f_0 .

Given the complexity of the Doppler term (due to the cosine in the argument of the exponential), one can express it in a more suitable expression using the Jacobi-Anger expansion

$$e^{i\Phi_D(t)} = \sum_{k=-\infty}^{+\infty} b_k e^{2\pi i k f_m t}, \quad (21)$$

where

$$b_k = J_k(\beta) e^{ik(\frac{\pi}{2} - \phi)}, \quad (22)$$

where $\beta = 2\pi f_0 R_\oplus \sin(\theta)$ and $J_k(\beta)$ is a Bessel function of the first kind. This expansion states that the Doppler modulation can be seen as a harmonic function of the modulation frequency. In practice, this sum can be reduced to a finite range by only selecting the significant frequency bins contained within the bandwidth of the signal $B = (1 + \beta)$.

Once again, a convolution is needed in order to compute the Fourier coefficients associated with \tilde{b}_n

$$\tilde{b}_n = \sum_{k=-\infty}^{\infty} b_k \text{sinc}(\pi(k f_m T_{obs} - n)) e^{i\pi(k f_m T_{obs} - n)}. \quad (23)$$

Putting everything together, we can now write an analytic expression for the Fourier domain response as

$$\tilde{s}_n(f) = \frac{1}{2} e^{i\varphi_0} \sum_k \tilde{a}_k \sum_l \tilde{b}_l \sum_m \left(A_+ \tilde{p}_m^+ + e^{i3\pi/2} A_\times \tilde{p}_m^\times \right), \quad (24)$$

where $n = k + l + m$. Regarding the ranges of the sum coming from Eqs. (19), (20) and (22), only a handful of terms have a significant contribution to the total power of the signal. In fact, by selecting only the terms such that the argument of the cardinal sine is comprised between $\pm 18\pi$, we obtain 99% of the power of the original time domain response [8]. This puts restrictions on the sums above of $-18 \leq k - \text{int}(f_0 T_{obs}) \leq 18$, $-(1 + \beta) f_m T_{obs} - 18 \leq l \leq (1 + \beta) f_m T_{obs} + 18$ and $-4 f_m T_{obs} - 18 \leq m \leq 4 f_m T_{obs} + 18$.

A comparison of this method with the computation in the time domain described in Sec. II B, revealed that the waveform generation computation time can be decreased by a factor of as high as 50. Moreover, the match between the time and Fourier domain waveforms are always higher than 99%.

III. EVOLUTIONARY ALGORITHMS

Evolutionary algorithms (EAs) refer to a group of stochastic, population-based algorithms that mimic biological evolution and social behavior to solve global optimization problems. An advantage of these algorithms is that they function with no *a priori* knowledge of the

problem at hand. This allows, in most cases, a very easy implementation of the algorithm. Generally, EAs display good convergence properties and have a small number of control parameters.

Previous studies in GW astronomy have used EAs, but have focused on the implementation of a single algorithm per source type [12, 40]. Our experience has shown that a possible better strategy is to use a combination of algorithms in the development of efficient, accurate search and resolution pipelines. This idea of creating hybrid EAs is not new, and has already been applied in a number of fields, including GW astronomy [28, 41].

In this work, our goal is to construct a hybrid search algorithm, which predominantly uses a combination of PSO [23, 24] and DE [25] to iteratively search for monochromatic galactic binaries. These two methods will be supported by a number of other techniques that will accelerate the convergence of the algorithm. In order to both extract the best-fit parameters at the end of the search phase, and conduct a full statistical analysis, we will use a combined Metropolis-Hastings - DE Markov Chain [42, 43].

In this section we describe all algorithms in their base form, before moving on, in the next section, to describing the full construction of our algorithm.

A. PSO

PSO is an evolutionary algorithm that uses swarm dynamics in order to solve optimization problems. This algorithm has already been used in other fields of astrophysics, such as pulsar timing [44, 45], ground-based GW astronomy [46] and cosmic microwave background studies[29]. The PSO is particularly adapted to when trying to find the extremum of multimodal and non-trivial likelihood surfaces. As far as we know, this method has never been tested in the search for galactic binaries using a space based observatory.

As mentioned before, the PSO mimics swarm dynamics in nature to find the extremum of a surface. This surface is parametrized by the value of a so-called fitness function that depends on a number of model parameters. In our case, we equate the fitness function to the likelihood $\mathcal{L}(\lambda^\mu)$. The motion of the swarm of particles on the parameter surface is parametrized by a fictitious time parameter t_j , where j is the identity of the current step. Each particle is then evolved via the standard dynamical PSO equations:

$$X^i(t_{j+1}) = X^i(t_j) + V^i(t_j), \quad (25)$$

$$V^i(t_{j+1}) = wV^i(t_j) + c_1\xi_1(P^i(t_j) - X^i(t_j)) + c_2\xi_2(G(t_j) - X^i(t_j)), \quad (26)$$

where the dynamical variables for particle i are the instantaneous coordinate position $X^i(t_j) = \{\lambda_i^\mu\}$ and velocity $V^i(t_j)$. At each position of the particle on the

parameter surface, we associate the corresponding value of the likelihood $\mathcal{L}(X^i(t_j))$.

The velocity of the particle involves a number of quantities specific to PSO. A further specificity of the PSO algorithm is that each particle retains a partial memory of aspects of their personal history, while the swarm as a whole retains a sense of global history, both of which effect the velocity evolution of the particle. With this in mind, the first important quantities to define are the notions of the personal best, P_{best}^i , and group best, G_{best} , positions.

We define $P_{best}^i(t_j)$ to be the maximum value of the fitness function for particle i during the duration of its history; i.e.,

$$P_{best}^i(t_j) = \max [\mathcal{L}(X^i(t_k)) \text{ , for } k = 0 \dots j] . \quad (27)$$

The personal best position P^i of a particle i at time t_j is then defined to be the position where the particle likelihood was equal to P_{best}^i

$$P^i(t_j) = X^i(t) \text{ if } \mathcal{L}(X^i(t)) = P_{best}^i(t_j). \quad (28)$$

Equivalently, we define $G_{best}(t_j)$ to be the maximum value of the fitness function of the swarm over its history, or, in other words, the maximum among all P_{best}^i ; i.e.,

$$G_{best}(t_j) = \max [P_{best}^i(t_j) \text{ , for } i = 0 \dots N_p] , \quad (29)$$

where N_p is the total number of particles in the swarm. The group best position is then obtained via

$$G(t_j) = P^i(t_j) \text{ if } P_{best}^i(t_j) = G_{best}(t_j). \quad (30)$$

Thus, at any time, each particle of the swarm has a personal memory through P_{best}^i and a group memory through G_{best} .

The remaining terms in Eq. (26) are an inertia w , two acceleration constants (c_1, c_2), and two scaling factors (ξ_1, ξ_2). For these parameters, the standard values used in the literature are $w = 0.78$, $c_1 = c_2 = 1.192$ and $(\xi_1, \xi_2) \in U[0, 1]$ [47]. This choice of parameters is believed to provide a good compromise between exploration and convergence for the swarm in a reasonable number of steps. However, as we will see later, these values can be made to take different values during the evolution of the algorithm.

On further investigation of Eq. (26), we see that there are three distinct contributions to the evolution of the velocity at each step :

- an inertial term that scales the current velocity with a factor w . If $w > 1$, the velocity will increase at each step and the particles are more likely to explore the parameter space, while for $w < 1$, the particles tend to converge to a single point in parameter space.
- an acceleration term toward the best position of the particle so far P^i . This term tends to make the particle explore the neighbourhood of a promising location in parameter space.

- an acceleration term toward the best position of the swarm so far G . This term is shared between all of the velocity equations of the swarm. This is the key term that creates the swarm dynamics: particles will tend to help each other to get to the best position of the group.

Since the motion of the particles is governed only by the dynamical equations given above, it is essential to set boundary conditions so that the position of the particles stays within the physical parameter range of interest. One of the common boundary conditions used for PSO is the reflective boundary conditions, where if the particle crosses the physical boundary, the position of the particle is set at the boundary while the sign of its velocity is reversed; i.e.,

$$\begin{aligned} X^i(t_{j+1}) &= X_{min} \text{ or } X_{max}, \\ V^i(t_{j+1}) &= -V^i(t_j). \end{aligned} \quad (31)$$

Other boundary conditions can be implemented depending on the nature of the model parameter. We give more details on the boundary conditions we have applied in the next section.

Finally, we also set limits on the values that the velocity can take in order to prevent the particle from moving outside the physical boundary of the problem. If the velocity computed at time t_{j+1} is superior to a set maximum velocity V_{max} , the velocity is then set at this limit; i.e.,

$$V^i(t_{j+1}) = \begin{cases} V_{max} & \text{if } V^i(t_{j+1}) \geq V_{max} \\ -V_{max} & \text{if } V^i(t_{j+1}) \leq -V_{max} \end{cases} \quad (32)$$

where $V_{max} = \frac{1}{4}(X_{max} - X_{min})$, X_{min} and X_{max} being the superior and inferior limit, respectively, for the given parameter.

As PSO was originally formulated to mimic flocks of birds, there is no real evolution as the population moves as a whole. At each point in time, the position of a particle is influenced by all other members of the swarm. This is contrast to most EAs. As we have seen above, the PSO is quite a simple algorithm as it is dependent on only three control parameters (w, c_1, c_2). The final dependency is the number of particles in the swarm, N_p . However, a downside of PSO is that there is no guide to choosing the optimum value of N_p . As we will see later, one needs to be careful when choosing N_p in order to achieve a balance between accuracy and runtime of the algorithm.

B. DE

In this work, DE will be used in two different and distinct manners. In the first instance, a pure form of DE is used as part of the search phase of the pipeline. In the second, it is used as an MCMC variant during the parameter estimation phase.

DE works by evolving a population of N_p candidate solutions in the parameter space. Each member of the population evolves via the creation of new candidate solutions by combining existing solutions into a mutant solution. This mutation then survives depending on the evaluation of a fitness function (we again use the likelihood function). A major advantage of DE is that, once again, it is very capable of handling multimodal solutions.

In this study, we evolve the population via a simplified DE rule : starting with a population of N_p particles at generation g , a mutated solution is produced at generation $g + 1$ according to

$$\bar{X}^i(g + 1) = X^j(g) + \gamma [X^k(g) - X^l(g)], \quad (33)$$

where $i \neq j \neq k \neq l \in [1, \dots, N_p]$. This means that a minimum of four particles is needed for DE to work. The differential weight $\gamma \in (0, 2]$ is a real, constant factor that controls the amplification of the differential vector $[X^k(g) - X^l(g)]$. Using this mutant solution, a crossover solution is created by combining elements of the target solution $X^i(g)$, with the mutated solution $\bar{X}^i(g + 1)$, to produce a trial solution $X^i(g + 1)$. In general, this solution is then evaluated via a greedy criterion of if $\mathcal{L}(X^i(g + 1)) > \mathcal{L}(X^i(g))$, then accept the new solution.

In our implementation of the algorithm, we neglect the crossover step, and for reasons that we justify later, we also replace the greedy criterion for acceptance with an evaluation of the Metropolis ratio,

$$H_M = \frac{\mathcal{L}(X^i(g + 1))}{\mathcal{L}(X^i(g))}, \quad (34)$$

where the new solution is accepted with a probability of $\alpha = \min(1, H_M)$. As with PSO, the performance of DE is controlled by the values of the differential weight γ , the number of particles N_p , and the crossover probability.

C. Metropolis-Hastings Markov Chain

The Metropolis-Hastings variant of the Markov Chain Monte Carlo (MCMC) algorithm [26, 27] is quite commonly used in the field of Bayesian inference. This quite simple algorithm works as follows : starting with a signal $s(t)$ and some initial template $h(t; \lambda^\mu)$, we choose a starting point $x(\lambda^\mu)$ randomly within a region of the parameter space, bounded by the prior probabilities $\pi(\lambda^\mu)$, which we assume to contain the true solution. We then draw from a proposal distribution and propose a jump to another point in the space $x'(\lambda^\nu)$. To compare the performance of both points, we evaluate the Metropolis-Hastings ratio

$$H = \frac{\pi(x')p(s|x')q(x|x')}{\pi(x)p(s|x)q(x'|x)}. \quad (35)$$

Here $q(x|x')$ is a transition kernel that defines the jump proposal distribution, and all other quantities are previously defined. This jump is then accepted with probability $\alpha = \min(1, H)$; otherwise the chain stays at $x(\lambda^\mu)$.

In order to improve the overall acceptance rate, the most efficient proposal distribution to use for jumps in the parameter space is, in general, a multivariate Gaussian distribution. While we have stated previously that the Fisher information matrix (FIM) is not a good choice tool for parameter estimation studies, it is an acceptable first-order approximation to the true distribution (i.e., the general directionality of the eigenvalues and the scale of the eigenvectors coming from the FIM produce acceptance rates that allow the algorithm to converge within a reasonable timescale). We define the FIM as

$$\Gamma_{\mu\nu} = \left\langle \frac{\partial h}{\partial \lambda^\mu} \middle| \frac{\partial h}{\partial \lambda^\nu} \right\rangle = -E \left[\frac{\partial^2 \ln \mathcal{L}}{\partial \lambda^\mu \partial \lambda^\nu} \right], \quad (36)$$

where the second term relates the FIM to the negative expectation value of the Hessian of the log-likelihood, and can be interpreted as a local approximation to the curvature of the likelihood surface. For the multivariate jumps we use a product of normal distributions in each eigendirection of $\Gamma_{\mu\nu}$. The standard deviation in each eigendirection is given by $\sigma_\mu = 1/\sqrt{DE_\mu}$, where D is the dimensionality of the search space (in this case $D = 7$), E_μ is the corresponding eigenvalue of $\Gamma_{\mu\nu}$ and the factor of $1/\sqrt{D}$ ensures an average jump of $\sim 1\sigma$. In this case, the jump in each parameter is given by $\delta\lambda^\mu = \mathcal{N}(0, 1)\sigma_\mu$. This type of MCMC algorithm is commonly referred to as a Hessian MCMC.

While the pure form of DE is used during the search phase, we can also use a variant of DE in the parameter estimation phase. By combining DE with a Metropolis-Hastings MCMC, commonly known as Differential Evolution Markov Chain (DEMC) [42, 43], we can accelerate the convergence of the MCMC. In practice, N_c simultaneous chains are used to explore the posterior density. For chain i at iteration l , the next iteration is proposed via

$$X_{l+1}^i = X_l^i + \gamma(X_l^j - X_l^k), \quad (37)$$

where $i \neq j \neq k \in [1, \dots, N_c]$ and the differential weight has the optimized value of $\gamma = 2.38/\sqrt{2D}$ [42, 43], with D being the dimension of the search space parameter. We note here that we also use this value in the standard implementation of the DE during the search phase.

Again, the number of chains N_c is problem dependent. While a larger number of chains creates a larger well from which to draw the next proposal, it also creates a computational bottleneck. As with previous works [35], we use a trimmed single chain history to circumvent this problem. This now means that $i \neq j \neq k \in [1, \dots, N_{TH}]$, where N_{TH} are the number of trimmed history points. We have found that $N_{TH} = N_{MCMC}/10$, where N_{MCMC} is the total number of iterations in the DEMC, provides a good enough history from which to draw points.

IV. BUILDING A HYBRID SWARM ALGORITHM

In this section, we now present how the previous algorithms can be combined in the framework of the detection and resolution of galactic binaries.

A. Defining the GB parameter space

In Sec. II B, we have seen that the gravitational waveform of a circular monochromatic binary is parametrized by a set of seven parameters. This sets the dimensionality of the monochromatic GB problem at $7 \times N$, where N is the total number of ultra compact binary sources in the Galaxy.

In order to reduce the dimensionality of the search space, one can apply an analytical method called the F-statistic [38]. The idea is to split the seven parameters in two distinct sets labeled extrinsic and intrinsic. The extrinsic parameters, (i, ψ, A, ϕ_0) , are related to the detector position and the orientation of the source. The intrinsic parameters, (f_0, θ, ϕ) , are dependent on the dynamics of the astrophysical source (note that θ and ϕ are intrinsic because of the movement of the detector with respect to the source). One can then show that for any set of intrinsic parameters, the likelihood can be equated to the F-Statistic in a form that automatically maximizes over the extrinsic parameters. Using the F-statistic reduces the dimensionality of the galactic binary search space from $7 \times N$ to $3 \times N$. Full details regarding the F-statistic can be found in Appendix A.

While the F-Statistic is very useful in finding the maximum of the likelihood surface, it should not be used when mapping the posterior density. In fact, the maximization of the extrinsic parameters prevents proper exploration of the posterior density and, as a consequence, parameter estimation can be incorrect [6]. However, unless otherwise stated, we will use the F-Statistic with all algorithms in the search phase of the pipeline.

B. Setting a detection threshold

To set a SNR detection threshold for eLISA, we ran a series of null tests. In this case, we assume that the output of the detector is composed of instrumental noise only, i.e. $s(t) = n(t)$. It has been shown that previous null tests for inspiralling supermassive black hole binaries in eLISA provide a detection threshold of $\rho = 10$ [48].

In this case we search the detector output using monochromatic waveforms. To conduct the null test, we used a modified DEMC algorithm. To encourage the movement of the chain we use a combination of thermostated and simulated annealing [16, 17, 49]. This replaces the factor of $1/2$ in the likelihood with an inverse temperature $\gamma = 1/(2T)$. The thermostated annealing,

defined by

$$\gamma = \begin{cases} \frac{1}{2} & 0 \leq \rho \leq \rho_0 \\ \frac{1}{2} \left(\frac{\rho}{\rho_0} \right)^{-2} & \rho > \rho_0 \end{cases}, \quad (38)$$

allows the algorithm control over how much heat needs to be injected into the likelihood surface. In this case, we took $\rho_0 = 1$ as we wanted the chain to begin exploring as quickly as possible. To extract the threshold SNR, we then need to cool the surface down slowly. The simulated annealing phase, defined by

$$\gamma = \begin{cases} \frac{1}{2} 10^{-\xi \left(1 - \frac{i}{t_{cool}} \right)} & 0 \leq i \leq t_{cool} \\ \frac{1}{2} & i > t_{cool} \end{cases}, \quad (39)$$

is then used to cool the surface. In the above expression, $\xi = \log_{10}(T_{th})$, where T_{th} is the temperature at the end of the thermostated annealing phase, i is the iteration number, and t_{cool} is the cooling schedule for the simulated annealing. For the null tests, we used 5×10^4 iterations each of thermostated annealing, simulated annealing, and standard DEMC.

We ran the above algorithm 50 times, with different starting configurations. In each case, the algorithms returned “detections” with SNRs of $5.9 \leq \rho \leq 8.7$. To account for the possibility of higher values from different noise realizations, we thus decided to set the detection threshold at $\rho = 9$.

C. Single-source detection

Our initial goal was to investigate and define regions of functionality for each component of the search pipeline. This top-down approach would allow us to confidently detect a source within a small region of parameter space surrounding the binary, and then expand the size of the search space, solving problems as we go. With this in mind, we defined a search region around the binary in each case by $\lambda^i \pm n\Delta\lambda^i$, where $n > 1$. In our initial investigations, we defined $\Delta\lambda^i = \sigma_i$, where σ_i is the standard deviation calculated using the FIM (we point out again that we are not using the FIM for parameter estimation purposes, but only to define a region of search space). In later investigations where we are dealing with a wider search space, for frequency, we take $\Delta f_0 = f_m = 1/yr$.

We used two criteria to assess whether the algorithm was working for a given width n : does it recover the SNR of the source, and are the recovered parameters less than 3σ away from the true solution? If those two conditions were met for repeated simulations, we increased the width of the search space. Thus, we were able to control both how our algorithm works and observe exactly when it breaks down.

In order to test the performance of the algorithms, we started with a simple search where the data contained a

single binary and an observation time of one year. Two different sources were selected for this first test phase. The first one was a fiducial low-frequency source, which allowed us to reduce runtime of the algorithm during development and solve problems in our initial investigations quickly. The parameters for this binary were $\iota = 1.74$, $\phi_0 = 1.664$, $A = 1.2276 \times 10^{-20}$, $\psi = 1.884$, $f_0 = 5.2234 \times 10^{-4}$ Hz, $\theta = 1.273$, and $\phi = 1.8845$, where the angular values are in radians. The signal had an SNR of $\rho \approx 54$. The second source was the WD-WD GB, RXJ0806.3+1527. This source is one of the main verification binary candidates for eLISA and should be the one recovered with the highest SNR [5]. The parameters for this binary are $\iota = 0.663$, $\phi_0 = 5.857$, $A = 6.378 \times 10^{-23}$, $\psi = 1.741$, $f_0 = 6.2203 \times 10^{-3}$ Hz, $\theta = 1.162$, and $\phi = 3.612$ [50]. For 1 year of data, the SNR for RXJ is $\rho \approx 25$

1. Initial search with PSO

The initial step in the construction of our algorithm was to implement the PSO to see how well it performs, on its own, in the detection of galactic binaries. There were two motivations to use PSO as a baseline algorithm :

- In comparison with MCMC based algorithms, the performances of PSO seemed to be more efficient in finding the maximum of complicated posteriors [29].
- The swarm-like feature of the algorithm, such as the personal and group best position, seemed to be powerful and could easily be adapted, or complemented, in other situations.

Since the PSO algorithm described in Sec. III A is very general, we need to explain some specificities for the GB search. The set of intrinsic parameters used for a single source PSO search is $\lambda^\mu = \{\ln f_0, \theta, \phi\}$. We do not use the cosine of the colatitude in this case, as we want the particles to evolve freely on the 2-sphere of the sky. In terms of boundary conditions for the angles, we let the particles evolve freely in absolute values of the sky angles. If the value of the sky position goes outside the physical boundary, we coherently remap the source into the natural boundaries $\theta \in [0, \pi]$ and $\phi \in [0, 2\pi]$. For the frequency, we used the reflective boundary condition given in Eq.(31).

For the first application of PSO, we focused solely on the low frequency source. We initially used 10 particles with standard control parameter values of $w = 0.72$, $c_1 = c_2 = 1.192$ for a total number of steps of 500. Using $\Delta\lambda^i = \sigma_i$, for every chosen value of width n , we launched ten simulations with different initial conditions. In this configuration, the PSO works well up to $n = 10^2$, where one of the simulations did not converge to the true solution. This value of n corresponds to a $\sim 3f_m$ frequency

bandwidth, where some of the secondary peaks in the likelihood, caused by the Doppler modulation, are now accessible to the particles. In the failed simulation, the swarm was not able to escape a secondary mode in the likelihood surface.

One possible solution for this problem is to increase the number of particles, N_p , thus giving the swarm more opportunities to find the main mode. This solution is not costless since it also increases the total computation time. This is why we decided to keep this as a last-resort solution, and instead introduce a new control on the swarm. In order to fine-tune the behavior of the swarm we now allow the constant inertia w to vary over time, i.e. $w = w(t_j)$. In the dynamical equations, a value of inertia superior to 1 increases the exploration of the swarm, while convergence of the swarm is improved for inertias inferior to 1. A good compromise between exploration and convergence can be found by introducing a type of “inertia annealing”. This method is not new and has already applied in other works, albeit in a different form [45]. Starting with an initial inertia w_i , we cool the inertia according to

$$w(i) = \begin{cases} w_f 10^{\log_{10}(\frac{w_i}{w_f})(1-\frac{i}{T_w})} & \text{if } 0 \leq i \leq T_w \\ w_f & \text{if } i > T_w, \end{cases} \quad (40)$$

where $w_i = 1.2$, the final inertia is $w_f = 0.78$ and $T_w = 500$ is the cooling time (which we take to be equal to the total number of steps in the algorithm). With this new annealed version of the algorithm, the swarms now successfully found the source up to a value of $n = 10^4$. For this value of n , the search space now covers the full sky and is $\sim 300f_m$ wide in frequency.

Beyond this value of n , we seemed to come to a natural limit in the PSO algorithm. The efficiency of the PSO is based on the ability of members of the swarm to find good positions in parameter space, and drag the whole set of particles toward them. One of the fundamental requirements is then to have a good exploration of the parameter space so that the swarm is not confined in some small part of the likelihood surface. For medium size widths such as the one with $n = 10^4$, we noticed that for some initial conditions, the swarm was not able to explore a large enough space to find the area of interests. It is known that PSO can very quickly converge to secondary maxima, especially on a complicated surface. Once there, it becomes very difficult to move the algorithm on to a better solution. This is why we decided to introduce a DE step that would provide the swarm a greater ability to explore the parameter space.

2. Combining PSO with DE

In the DE part of the combined PSO-DE algorithm, all of the particles are evolved sequentially using Eq. (37). This is considered to be one step in the pipeline. At the

end of the step, we update the values of the personal and group best positions of the swarm. Because of its different nature, DE is more suited for global exploration and large moves in the search space. Therefore, the main idea was to alternate between standard PSO blocks as described before and DE blocks that would help find better P^i and G for the next stage of PSO. From the beginning, we also decided to use the thermostated and simulated annealing schemes for the DE.

We implemented our algorithm using two series of PSO and DE blocks of 125 steps each for a total number of 500 steps. As we have already described, both the PSO and the DE have their own specific annealing schemes. As we alternate between the PSO and the DE, we also need to alternate between the annealing programs. To ensure that the use of these schemes was optimal, they were implemented in the following manner : both the inertia annealing (for the PSO) and the thermostated/simulated annealing (for the DE) were run simultaneously. This ensured that each annealing scheme cooled over a long period. It also meant that when one scheme was in use, the other was running in the background, ensuring that when it came back into use, it would be at the same level as if it was being applied to a single algorithm pipeline only. So, for the PSO, we set $T_w = 500$. This means that at the end of the algorithm, the inertia is equal to the final value of $w_f = 0.78$, even though the actual total number PSO steps is 250. During the DE phase, we used a thermostated annealing scheme during the first 125 steps, with a threshold of $\rho_0 = 5$ to encourage movement in the parameter space. We then set $T_c = 375$ for the simulated annealing phase, meaning that at the end of the pipeline, the heat is unity. In this new combined configuration, the algorithm managed to always find the source for $n = 10^5$, which corresponds to a frequency bandwidth of $\sim 3000f_m \approx 10^{-4}\text{Hz}$

At this point, we chose to test the algorithm on a full 1mHz frequency band to observe its performance. The interest here is to see whether, given the width of the signal in comparison to the width of the search space, the algorithm would find the source, or repeatedly get stuck on strong peaks in the noise, i.e., a kind of “needle in an empty field” problem. Since the frequency band is now much wider, we decided to keep the structure of the previous algorithm, but it quickly became clear that we finally needed to increase the number of particles in the swarm.

We thus changed the number of particles from 15 to 40 to facilitate a wider exploration (we comment further on the number particles required below). In this configuration, the algorithm was doing quite well, but was not able to satisfy the two convergence criteria for all the simulations.

In fact, we observed two main weaknesses in the movement of the swarm: in the case where the source has not yet been detected, the group best position G may be lying on a distant noise peak. A particular particle may land in the vicinity of the source, and in so doing, improve its

personal best position P^i . However, as $\mathcal{L}(P^i) < \mathcal{L}(G)$, the group best position will never move into the area of interest. Furthermore, while P^i stays close to the source, the particle itself wanders off to explore the rest of the parameter space. If a new personal best position is found, it can actually pull P^i away from the source, delaying detection. In the second case, a source can be detected (i.e., its SNR beats the threshold), but the group best position ends up on a secondary solution. In order to improve on this current solution, we have to wait until one of the particles lands at random on a better solution and thus improving $\mathcal{L}(G)$.

Upon investigation, the culprit in both cases was the width of the search window. A wide search window allows for a very diffuse swarm. This means that any chance of a local exploration is dramatically reduced. In this case, particles will visit the region of parameter space close to the source once, and then maybe never again after. One solution would be to overpopulate the search space with particles, increasing the possibility that certain regions would be visited by many particles, many times, during the runtime of the algorithm. However, this would lead to a very slow algorithm. So, in order to solve these two issues, while keeping the number of particles small, we introduced two further features to the algorithm that enforce a more local exploration and reduces the width of the search space.

3. Uphill Climber

The motivation for this type of move is fact that the swarm algorithm's lack of local exploration possibilities. In the previous test cases, the frequency band was small enough that the swarm was able to locally explore good positions on the likelihood surface. For wider frequency bands, the swarm is more diffuse, and the PSO-DE blocks do not provide an opportunity for the swarm to locally explore the personal best positions P^i .

The idea was then to implement a new block of MCMC-like movements that would be related, not to the individual swarm particles, but directly to the P^i . Thus, at some time t_j , we stop the global movement of the swarm and locally explore all P^i positions that have been spotted so far with the PSO and DE blocks.

The uphill climber (UC) scheme was first introduced in [41] as a greedy criterion proposal, i.e., if the new solution has a higher fitness than the starting solution, move there immediately. In the current context, this move is implemented as follows : for each $P^i(t_j)$, and as we are using the F-Statistic, calculate the FIM on the 3D-projected subspace at that point, i.e., iteratively project the D -dimensional FIM onto a $D - 1$ subspace according to

$$\Gamma_{\mu\nu}^{(D-1)} = \Gamma_{\mu\nu}^{(D)} - \frac{\Gamma_{\mu\kappa}^{(D)} \Gamma_{\nu\kappa}^{(D)}}{\Gamma_{\kappa\kappa}^{(D)}}. \quad (41)$$

As with the MCMC algorithm, use the eigenvalues and eigenvectors of the projected matrix to construct jump

proposals, and accept or reject according to the greedy criterion $\mathcal{L}(\lambda^{new}) > \mathcal{L}(\lambda^{old})$.

In order for the UC to be effective, it needs to be implemented a number of times successively as the acceptance rate is usually quite low ($< 10\%$). Although low, and probably coming from the fact that we do not update the eigenvectors and eigenvalues after a move has been accepted, if the UC is used N_{UC} times in sequence (with $N_{UC} \geq 10^2$), it can easily move the position of P^i between $10 - 30 f_m$ in frequency. This greatly accelerates the convergence of the algorithm. To make the UC fit into the overall structure of the pipeline, and especially with the annealing schemes, we count the N_{UC} iterations as one step in the algorithm. This means that we take the thermostated/simulated annealing temperature at t_j and keep it constant during the UC moves.

4. Swarm Strengthening and Culling

The second feature was introduced due to the fact that the likelihood surface has many secondary peaks because of the multi-modal nature of the detector response. One of the strongest secondary solutions comes from the Doppler modulation of the phase. For some simulations, the swarm was able to get close to the source but was stuck on one of these Doppler induced peaks. As mentioned above, when the frequency band is large, even with a local exploration of the likelihood surface, only a handful of particles explore the area of interest, and those that do so may converge to one of the secondary solutions. In order to prevent this situation, we decided to add a second phase in the overall algorithm where we strengthen a good solution by moving half of the swarm to an interval of frequency around G given by $[f_0(G) \pm 10 f_m]$. The size of this band is motivated by the assumption that, if this feature is used late enough in the pipeline, the principal mode should not be further away than a few f_m from the secondary solutions.

Furthermore, while it is clear that an increased number of particles is necessary in the early stages of the pipeline to ensure a wide exploration of the parameter space, later on, some of these particles can converge to low fitness regions of the likelihood surface. Seeing as they never evolve to the region of interest during the runtime of the pipeline, we cull the 50% of the swarm that is not drawn to the region around G . This also helps speed up the runtime of the algorithm in the second phase of the pipeline.

5. Single source search over a 1-mHz band

As the primary goal of this study is accuracy, rather than speed, the final version of our algorithm is as follows : the pipeline is now 1250 steps long, beginning with 40 particles. In the first phase, the structure is two blocks of 150 steps each of PSO and DE, followed by 200 steps of

	ι	ϕ_0	A	ψ	f_0/mHz	θ	ϕ	SNR
Source 1	1.74	1.664	1.2275×10^{-20}	1.884	0.52234	1.273	1.8845	53.92
	1.736	1.707	1.2429×10^{-20}	0.3076	0.52233969	1.239	1.9097	53.95
RXJ0806.3	0.663	5.8565	6.378×10^{-23}	1.74	6.2202766	1.1624	3.6116	24.93
	1.143	1.785	9.769×10^{-23}	0.742	6.2202745	1.1600	3.6236	25.53

TABLE I: The injected and recovered parameter values for our two test sources. The values for RXJ0806.3+1527 are taken from [50]. We provide the recovered results for two simulations that end up on the response invariant symmetric solution $(\phi_0, \psi) \rightarrow (\phi_0 \pm \pi, \psi \pm \pi/2)$.

UC. For the PSO we use an inertia annealing scheme with $T_w = 750$, while for DE and UC, we use a thermostated annealing scheme. At the end of this phase, we move the swarm that is now 20 particles in size to the vicinity of G and kill the other 20 particles. In the second phase of the pipeline, we have a structure of 75 steps of PSO, followed by 75 steps of DE and 100 steps of UC. For DE and UC, we use a simulated annealing scheme with an initial temperature equal to the temperature at the end of the first phase and a cooling time of $T_c = 250$. We again point out that this structure is not optimized and will be a focus of our research in the future.

The search algorithm was then tested for the detection of the low frequency source, using a 1-mHz search band. We ran 10 simulations with different initial conditions, and in each case recovered the source according to both detection criteria. In Table I, we present both the injected and recovered parameter values from the search. For this source, the total runtime for the search algorithm was 1 minute on an Intel Xeon 2.6 GHz processor.

We then ran the algorithm on the more realistic source, namely the WD-WD verification binary RXJ0806.3+1527. Again, in all simulations, the algorithm was able to find the global solution. In Fig 1 we plot the evolution of the group best position G for the parameters (A, ι, f_0) , as well as the evolution of SNR, as a function of step number. We can see that the algorithm performs very well and we converge close to the true value of frequency in less than 10 steps of the algorithm. The sky angles take a lot longer to pin down, but are also eventually found. In the final cell of the figure we plot the evolution of SNR as a function of step number. We draw attention at this point to the values of the parameters and the SNR at 300, 600 and 750 steps in the plots. At each of these values we have an uphill climber move. We can see from looking at θ and ϕ that the UC moves are responsible for locking G onto the true values of the parameters. In the SNR plot, we can see that the UC produces large jumps in the recovered SNR. This again justifies the inclusion of this type of move. For this source, and due to the much higher sampling rates required, the total time for the search algorithm was 12.5 minutes. Again, we point out that there is a large potential for future optimization of the algorithm. We should also highlight here that for long periods of time in the plot, it looks like the algorithm has failed to move the solution and that it is static. We remind the reader again that the quantity being plotted is the group

best value G . An investigation of the particle positions displays a constant evolution of the swarm.

In Table I we again present both the injected and recovered parameter values from the search. We should point out that the recovered results presented are for two simulations that end up on a $(\phi_0, \psi) \rightarrow (\phi_0 \pm \pi, \psi \pm \pi/2)$ symmetric solution that leaves the response $s(t)$ invariant. This seems to be quite a common feature for a single channel observatory, and is a perfectly valid solution due to the multimodal nature of the response. In keeping with recent articles on supermassive black hole binaries [35, 51], we also carried out a full statistical analysis on the Markov chains for each source. For the low frequency source, all distributions had low values of skewness and kurtosis. However, for RXJ0806.3+1527, the posterior distributions were highly skewed and/or had a high kurtosis for a number of parameters. This demonstrates that the FIM should also be avoided as a parameter estimation tool for compact galactic binaries.

As a consequence, in our full analysis, we present 99% credible intervals, \mathcal{C} , for the parameter errors such that

$$\int_{\mathcal{C}} p(\lambda^\mu | s) d\lambda^\mu = 1 - \alpha, \quad (42)$$

where for a 99% BCI, $\alpha = 0.01$. This is a degree-of-belief statement that the probability of the true parameter value lying within the credible interval is 99%; i.e.,

$$\mathbb{P}(\lambda_{true}^\mu \in \mathcal{C} | s) = 0.99. \quad (43)$$

For the positional resolution of the source, we can define an error box in the sky according to [36]

$$\Delta\Omega = 2\pi\sqrt{\Sigma^{\theta\theta}\Sigma^{\phi\phi} - (\Sigma^{\theta\phi})^2}, \quad (44)$$

where

$$\Sigma^{\theta\theta} = \langle \Delta \cos \theta \Delta \cos \theta \rangle, \quad (45)$$

$$\Sigma^{\phi\phi} = \langle \Delta \phi \Delta \phi \rangle, \quad (46)$$

$$\Sigma^{\theta\phi} = \langle \Delta \cos \theta \Delta \phi \rangle, \quad (47)$$

and $\Sigma^{\mu\nu} = \langle \Delta \lambda^\mu \Delta \lambda^\nu \rangle$ are elements of the variance-covariance matrix, calculated directly from the DEMC chains themselves. As well as the sky error box, we also calculate the orthodromic distance between the true and recovered sky positions. This is given by the Vincenty formula

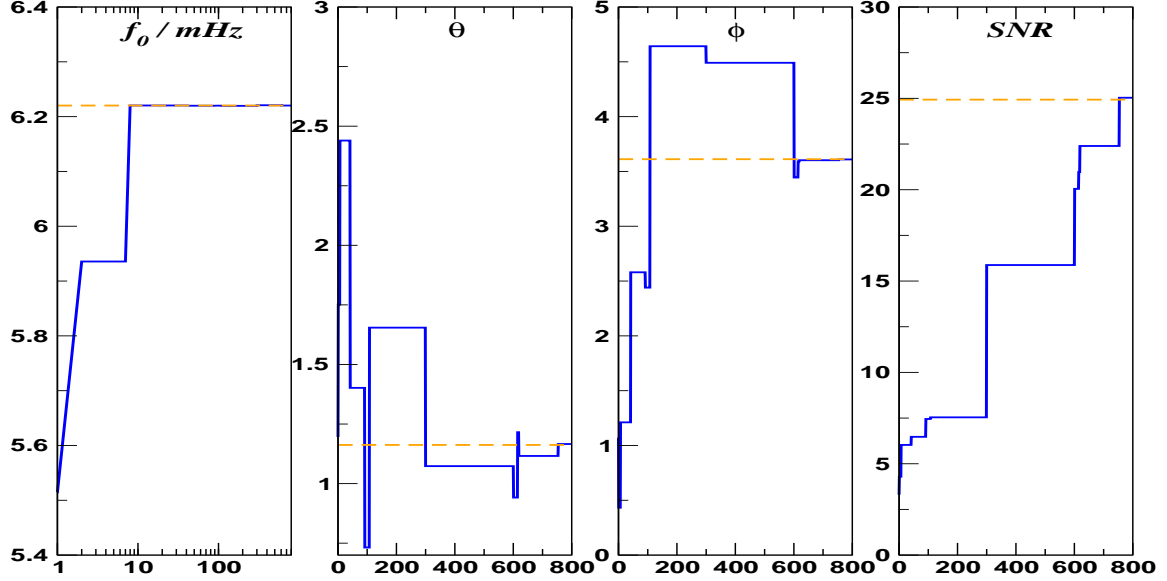


FIG. 1: A plot of the movement of the group best position G during the search for RXJ0806.3+1527, for the parameters (f_0, θ, ϕ) as a function of the number of steps in the algorithm. In each cell the true values are represented by the dashed (orange) lines. We can see that G converges to the true frequency in less than ten steps, whereas the sky angles take longer to find. In the final cell we plot the evolution of SNR. Here the large jumps in SNR at 300, 600 and 750 steps are due to the uphill climber move.

$$\Delta\sigma = \arctan \left(\frac{\sqrt{(\cos \phi_R \sin \Delta\theta_L)^2 + (\cos \phi_T \sin \phi_R - \sin \phi_T \cos \phi_R \cos \Delta\theta_L)^2}}{\sin \phi_T \sin \phi_R + \cos \phi_T \cos \phi_R \cos \Delta\theta_L} \right), \quad (48)$$

where we define the true longitude ϕ_T and latitude θ_T of the source, the recovered longitude ϕ_R and latitude θ_R , and $\Delta\theta_L = \theta_L^T - \theta_L^R$, and the value of $\Delta\sigma$ is in radians.

As the RXJ source is one of the main verification binaries of a space-based mission, we should comment here that for the eLISA configuration, with 1 year of data, the recovered frequency has an absolute error of 2 nHz or $0.06 f_m$, corresponding to a percentage error of $\sim 10^{-5}\%$. The recovered amplitude has an absolute error of $\sim 3 \times 10^{-23}$, while the absolute error in the measurement of inclination is approximately 29 degrees. And even though the sky position is known from EM observations, the sky error box based on 1 year of observation is $\Delta\Omega \approx 0.6 \text{ deg}^2$.

We should take some time here to talk about both the number of particles in the swarm, N_p , and the stopping criterion used for the pipeline. For the algorithm described above, we initially used 40 particles before culling the swarm to 20. In order to test for accelerated convergence, we ran the algorithm with different numbers of initial particles, up to $N_p = 10^2$. In no circumstance did we observe an acceleration of convergence that would convince us start with more than 40 particles. As ex-

pected, however, we did see an increase in the runtime of the pipeline that came without a corresponding increase in convergence or accuracy. As with most search algorithms, there is no predefined stopping criterion. This case is identical. An investigation of quantities such as P^i or G did not suggest any obvious way of using these parameters as stopping criteria. In this particular study, as we know the true values *a priori*, the number of steps in the algorithm was chosen to ensure that we always found the true source. We intend to investigate a more rigorous stopping criterion in a future work.

V. ANALYSIS AND RESULTS

A. Presentation of the problem

After the single source search, we decided to test our algorithm in a situation closer to reality where the data set contains several binaries. We have built two data sets with different attributes

- The first set of data contains 18 sources in a fre-

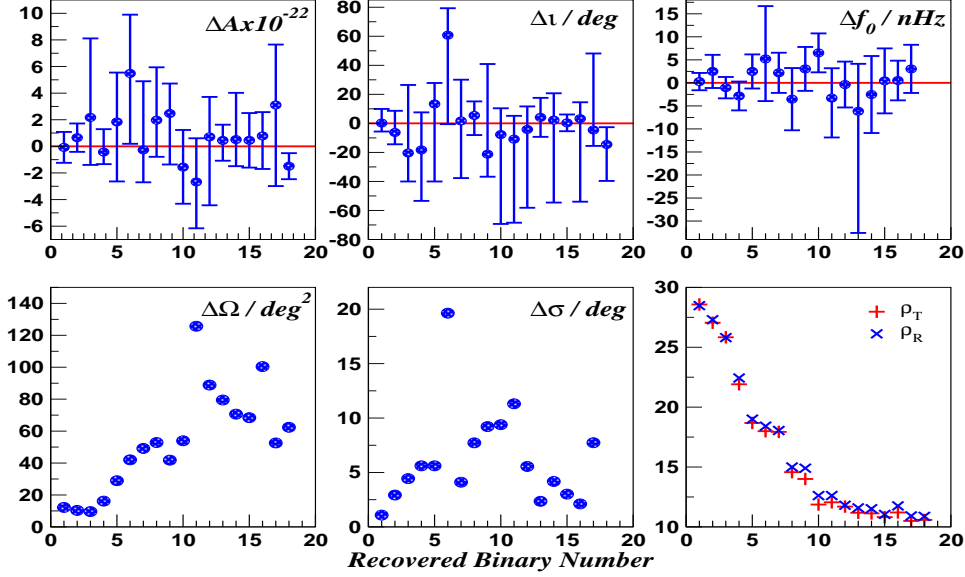


FIG. 2: Results obtained from the recovered sources in data set 1. In the top row, we present the 99% credible intervals and (true subtracted) median values for amplitude A (top left), inclination (top middle) and frequency (top right). In the bottom row, we present the values of the sky error boxes, the orthodromic distance, and the values of the true and recovered signal-to-noise ratios.

quency band of 1-mHz width from 0.5 to 1.5 mHz. The SNR of the sources have been taken between 10.5 and 28. The minimum distance in frequency between two sources is $273f_m$, meaning that there is no confusion between the sources, but there is plenty of opportunity for the algorithm to get stuck on a strong peak in the noise.

- The second data set contains 30 sources in a frequency band of $10^3 f_m \approx 30 \mu\text{Hz}$, centered at 1 mHz. The SNRs of the sources have been taken between 10.9 and 35. The minimum distance in frequency between two sources is $9f_m$, meaning that there is now a mild confusion between the sources.

At the end of the pipeline described above, we are faced with two practicalities that require different treatments. The first is that we need to have a way of quickly and accurately subtracting the recovered source before moving onto the search for the next source. The other is a full statistical analysis of the recovered source.

To complete the first task, at the end of the search phase, we run a 4×10^4 iteration DEMC with a constant inverse temperature of $\gamma = 1/2$. This gives the chain a chance to become statistically independent. We then run a further 2×10^4 iterations where we superfreeze the chain to extract the Maximum Likelihood Estimator (MLE) [6]. This is done by using a simulated annealing phase, where we use an inverse temperature of $\gamma = 1/(2T)$, starting with $T_i = 1$ and ending with $T_f = 10^{-5}$. At the same time, to conduct the statistical analysis and calculate the credible intervals, we run a 10^6 iteration DEMC, where we neglect the first 2×10^4 iterations as “burn-in”.

B. Data Set 1

For the first data set with no confusion, we detected 17 sources at the main frequency peak and 1 source on a secondary shifted frequency peak. In Fig. 2, we plot the (true subtracted) recovered median values for the parameters (A, ι, f_0) , i.e. $\Delta\lambda = \tilde{\lambda}_R - \lambda_T$, as well as the 99% credible intervals, as a function of the recovered binary. We also plot the sky error box $\Delta\Omega$, the orthodromic distance between the true and recovered sky positions $\Delta\sigma$, as well as the SNR.

For the amplitude and inclination, all the injected source values are contained in the credible intervals, except for source 6. For this source, the binary has almost no inclination. This leads to an extremely high anticorrelation between the parameters of -0.949. An inspection of the correlation matrices for the binaries using the information from the chains, we noticed that a high correlation value between amplitude and inclination leads to larger credible intervals. For this data set, we found nine binaries with correlations between A and ι superior to 0.9 in absolute value. In addition, most of the credible intervals for inclination are not symmetric around the recovered values, due to the high skewness and/or kurtosis of the posterior distributions. This confirms that credible intervals are indeed important to use in this situation. In terms of frequency, all the frequencies are recovered in the interval except for binary number 10. A closer inspection reveals that the median frequency from the chain is at 3.95σ away from true frequency, where σ is the standard frequency deviation from the chain. However, the

SNR of the recovered signal is the same as the source and the residual is below the noise. This means that we indeed have a detection. In the majority of cases, the recovered median frequencies were within 7 nHz, or $0.2 f_m$ of the true values. We should point out that source 18 is not represented in the frequency cell. We will tackle this binary separately below.

If we now focus on the bottom row of Fig. 2, we can see that, as expected, the sky error box tends to grow as a function of diminishing SNR. In almost all cases, the sky error box is smaller than 100 deg^2 , with some of the high SNR sources having error boxes of $\sim 10 \text{ deg}^2$. One binary (source 11) has a large error box ($\sim 150 \text{ deg}^2$). On inspection, we observed a large imprecision in the resolution of colatitude for this source, but the recovered value was within a 3σ of the true value. From the orthodromic distance, we can see that the recovered sky solutions are usually within $\sim 10 \text{ deg}$ of the true value. In the cell representing the SNRs, it is interesting to see that most of the binaries have been found in order of SNR. This essentially means that our search algorithm is able to locate the brightest signal on a 1-mHz band and finds the maximum of the fitness function in a noisy background.

The final source (#18), where we ended up on a secondary Doppler peak, was the binary with the lowest injected SNR of the set ($\rho_T = 10.58$). In all of our simulations, the MLE frequency extracted at the end of the search phase corresponded to a frequency that was $\sim 3f_m$ away from the true value. However, the recovered SNR from the MLE ($\rho_R = 11.08$) was actually higher than the injected value due to noise. It has been pointed out in the literature that this situation can happen, especially for sources with SNR close to the threshold [12]. As a sanity check, we ran 10 simulations with different noise realizations and tried to find this single source. For each simulation we managed to find the source on the true frequency peak, confirming that it was indeed a problem of noise realization, and not a fundamental problem with the algorithm.

To further test how we had done, we also computed the overlap between the true and recovered signals, where the overlap between h_T and h_R is given by

$$\mathcal{O} = \frac{\langle h_T | h_R \rangle}{\sqrt{\langle h_T | h_T \rangle \langle h_R | h_R \rangle}}. \quad (49)$$

The 17 fully recovered sources had overlaps above 0.9, with values as high as 0.999 for the highest SNR source. For source 18, the overlap was 0.82.

We can see what this means graphically by plotting the residual of the subtracted data set against the original signal and the instrumental noise. The residual power is an indication of how much we have disturbed the data set through the imperfect subtraction of a source. We define the residual as

$$\tilde{r}(f) = \tilde{n}(f) - \left(\tilde{s}(f) - \sum_{i=1}^{N_s} \tilde{h}_i^{MLE}(f) \right). \quad (50)$$

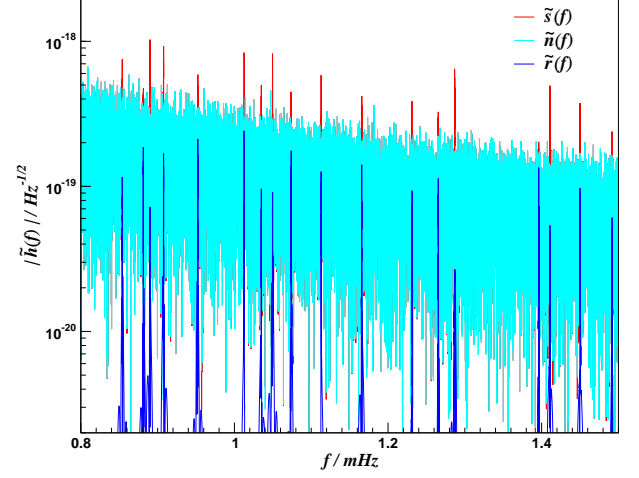


FIG. 3: A plot of the power spectra for the injected data set, the instrumental noise and the residual for data set 1.

with N_s the number of recovered sources. In Fig 3, we plot the power spectra for the total signal with noise $\tilde{s}(f)$, the instrumental noise only $\tilde{n}(f)$ and the residual $\tilde{r}(f)$. If the residual is below the level of the noise, then the subtraction process has been successful. For the first data set, we can see that the level of the residual is always below the level of noise even for the binary where the recovered frequency was $3f_m$ away.

We also carried out one final extra check to test the performance of the subtraction process. We ran 40 simulations, using different initial configurations, on the binary subtracted residual data set. If we had not properly extracted all binaries, we should in this case “detect” another source. However, all simulations ended up with a “detection” with an SNR inferior to the SNR threshold. Moreover, as some of the returned solutions were clustered around specific frequencies, we inspected these value and confirmed that they were pure noise peaks at more than $50f_m$ away from any injected frequency signals.

C. Data set 2

For the second data set, the algorithm managed to find all sources on the $10^3 f_m \approx 30 \mu\text{Hz}$ band. In Fig. 4, we again plot the credible intervals, sky errors, and SNR. For this data set, all the true values lie in the 99% credible intervals for amplitude, inclination, and frequency. Once again, we observe that the widths of the credible intervals for ι and A depend on the values of these parameters and their correlation. For instance, the largest credible interval for amplitude (binary 15) corresponds also to a high correlation between amplitude and inclination of 0.951 and a high asymmetry in the posterior density. In terms of frequency, the recovered values are very good and all lie at less than 3 nHz or $0.1 f_m$ from true frequency.

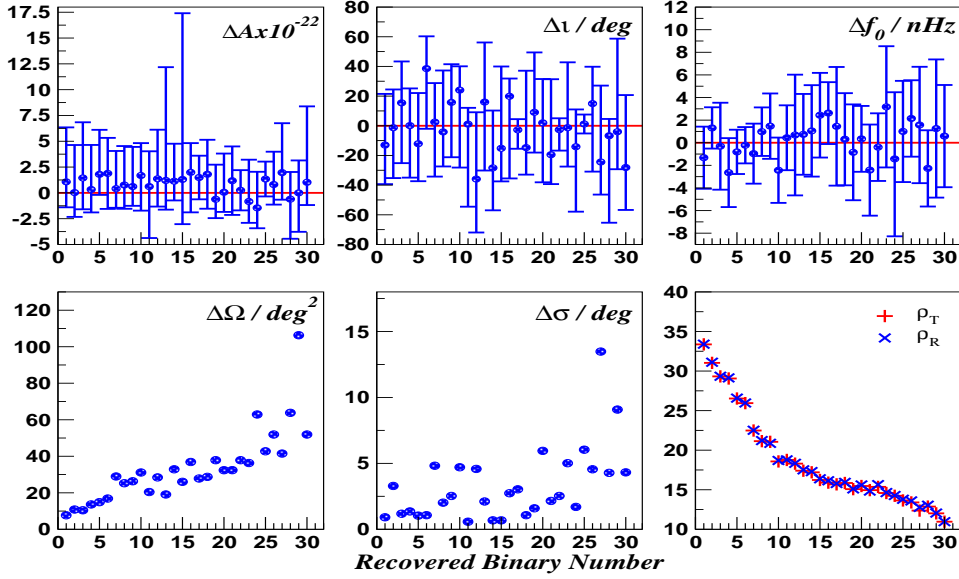


FIG. 4: Results obtained from the recovered sources in data set 2. In the top row, we present the 99% credible intervals and (true subtracted) median values for amplitude A (top left), inclination (top middle) and frequency (top right). In the bottom row, we present the values of the sky error boxes, the orthodromic distance and the values of the true and recovered SNRs.

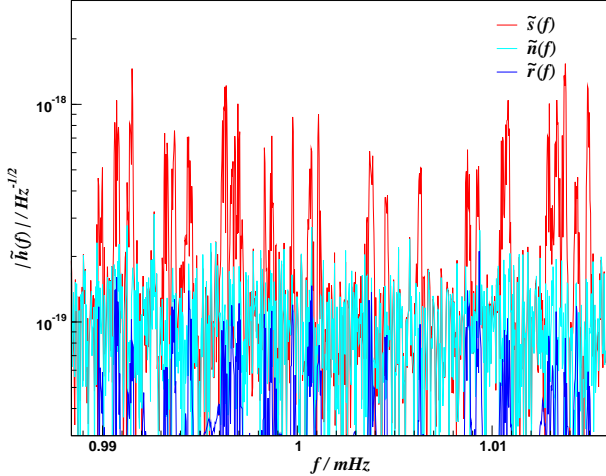


FIG. 5: A plot of the power spectra for the injected data set, the instrumental noise and the residual for data set 2.

For the sky positions, the sky error boxes have expected values. For this data set, the orthodromic distance between injected and recovered sky positions is lower, with nearly all sources recovered within 5 deg of the true value. Once again, the size of the sky error box increases as the SNR decreases, and the binaries we again found in order of decreasing SNR. In this case, and probably due to the smaller search band, the overlaps between the recovered and the true signals were once again extremely good and higher than 0.95 for all binaries.

In Fig. 5, we again see that the residual power is below the noise, suggesting a minimal disturbance of the data

set during source subtraction. And once again, running a sanity-test search on the source subtracted data set, we found that all simulations converged to a noise peak. Moreover, in this case, as the noise realization was different, all of our simulations converged to the same frequency peak, which returned an SNR value of $\rho = 5.52$, clearly below the threshold for detection.

VI. CONCLUSION

In this study, we have developed a hybrid swarm-based algorithm for the detection of gravitational waves emitted by ultra compact galactic binaries, mixing together Evolutionary Algorithms such as Particle Swarm Optimization and Differential Evolution. We demonstrated the ability of this algorithm to detect a single source on a 1-mHz band using a fiducial low frequency source and the verification binary RXJ0806.3+1527 in the framework of the future eLISA mission. We then applied this search algorithm to two data sets containing 18 and 30 sources on frequency bands equal to 1 mHz (no confusion) and $10^3 f_m \approx 30 \mu\text{Hz}$ (mild confusion) respectively. We successfully recovered all sources, and using a full Bayesian analysis, we demonstrated that the injected values were almost always within a 99% credible interval around the median values for the recovered binaries.

This demonstrated, that while unoptimized, the algorithm works well in the iterative search for GB sources. We have now started work on optimization of this version of the algorithm, and have also begun an investigation of the strong source confusion case, as well as an extension of the algorithm to the application of a simultaneous

multiple-swarm search.

Appendix A: F-statistic

The F-Statistic is a method for analytically maximizing over certain parameters in the GW response. As stated in the main body, for compact galactic binaries, we can split the set of waveform parameters into sets that we call intrinsic, i.e. $\{f_0, \theta, \phi\}$, and extrinsic, i.e. $\{\iota, \phi_0, A, \psi\}$. In the LFA, we can rewrite the detector response in Eq. (8) in such a way that it is expressed as a sum of constant amplitudes a_i , depending only on extrinsic parameters, and time varying functions $A^i(t)$, depending only on intrinsic parameters; i.e.,

$$h(t) = \sum_{i=1}^4 a_i A^i, \quad (\text{A1})$$

where we define the quantities a_i and A_i by

$$\begin{aligned} a_1 &= \frac{A}{2}((1 + \cos^2 i) \cos \varphi_0 \cos 2\psi - 2 \cos i \sin \varphi_0 \sin 2\psi), \\ a_2 &= -\frac{A}{2}(2 \cos i \sin \varphi_0 \cos 2\psi + (1 + \cos^2 i) \cos \varphi_0 \sin 2\psi), \\ a_3 &= -\frac{A}{2}(2 \cos i \cos \varphi_0 \sin 2\psi + (1 + \cos^2 i) \sin \varphi_0 \cos 2\psi), \\ a_4 &= \frac{A}{2}((1 + \cos^2 i) \sin \varphi_0 \sin 2\psi - 2 \cos i \cos \varphi_0 \cos 2\psi), \end{aligned} \quad (\text{A2})$$

and

$$\begin{aligned} A^1 &= D^+(t, \theta, \phi) \cos \Phi(t, f_0, \theta, \phi), \\ A^2 &= D^\times(t, \theta, \phi) \cos \Phi(t, f_0, \theta, \phi), \\ A^3 &= D^+(t, \theta, \phi) \sin \Phi(t, f_0, \theta, \phi), \\ A^4 &= D^\times(t, \theta, \phi) \sin \Phi(t, f_0, \theta, \phi). \end{aligned} \quad (\text{A3})$$

If we define a set of four constants $N^i = \langle s | A^i \rangle$, where s is the detector output, we can solve for the a_i 's according to

$$a_i = M_{ij}^{-1} N^j, \quad (\text{A4})$$

where the matrix M^{ij} is defined by

$$M^{ij} = \langle A^i | A^j \rangle. \quad (\text{A5})$$

If we now substitute these terms into the expression for the reduced log-likelihood

$$\ln \mathcal{L}(\lambda^\mu) = \langle s | h(\lambda^\mu) \rangle - \frac{1}{2} \langle h(\lambda^\mu) | h(\lambda^\mu) \rangle, \quad (\text{A6})$$

we can define the F-Statistic

$$\mathcal{F} = \log \mathcal{L} = \frac{1}{2} M_{ij}^{-1} N^i N^j, \quad (\text{A7})$$

which automatically maximizes over the extrinsic parameters.

Now, given the numerical values of the amplitudes a_i , we can solve for the extrinsic parameters according to

$$A = \frac{A_+ + \sqrt{A_+^2 - A_\times^2}}{2}, \quad (\text{A8})$$

$$\psi = \frac{1}{2} \arctan \left(\frac{A_+ a_4 - A_\times a_1}{-(A_\times a_2 + A_+ a_3)} \right), \quad (\text{A9})$$

$$i = \arccos \left(\frac{-A_\times}{A_+ + \sqrt{A_+^2 - A_\times^2}} \right), \quad (\text{A10})$$

$$\varphi_0 = \arctan \left(\frac{c(A_+ a_4 - A_\times a_1)}{-c(A_+ a_2 + A_\times a_3)} \right), \quad (\text{A11})$$

where

$$\begin{aligned} A_+ &= \sqrt{(a_1 + a_4)^2 + (a_2 - a_3)^2}, \\ &+ \sqrt{(a_1 - a_4)^2 + (a_2 + a_3)^2}, \end{aligned} \quad (\text{A12})$$

$$\begin{aligned} A_\times &= \sqrt{(a_1 + a_4)^2 + (a_2 - a_3)^2}, \\ &- \sqrt{(a_1 - a_4)^2 + (a_2 + a_3)^2}, \end{aligned} \quad (\text{A13})$$

$$c = \frac{\sin(2\psi)}{|\sin(2\psi)|}. \quad (\text{A14})$$

As with the waveform generation, one can optimize the computation time by evaluating the functions A^i directly in the Fourier domain. Once again the F-statistic computed directly in the Fourier domain is faster than that in the time domain by factors of up to 30.

Appendix B: Tables of recovered values

Here we present the injected and recovered values from the search pipeline for both data sets. For each binary, the top row represents the injected values, and the bottom row the recovered values.

-
- [1] G. Nelemans, L. R. Yungelson, and S. F. Portegies Zwart, *Astron. Astrophys.* **375**, 890 (2001).
 - [2] A. J. Ruiter, K. Belczynski, M. Benacquista, S. L. Larson, and G. Williams, *Astrophys.J.* **717**, 1006 (2010).
 - [3] G. Nelemans, *ASP Conf. Ser.* **467**, 27 (2013).

- [4] G. Nelemans, M. Wood, P. Groot, S. Anderson, K. Belczynski, M. Benacquista, P. Charles, A. Cumming, C. Deloye, P. Jonker, et al., White paper for the National Academics Astro 2010 (2010).
- [5] M. Kilic, W. R. Brown, and J. J. Hermes, in *9th LISA*

Binary	ι (rad)	ϕ_0 (rad)	A ($\times 10^{-22}$)	ψ (rad)	f_0 (mHz)	θ (rad)	ϕ (rad)	ρ
1	1.921	5.401	8.592	2.272	1.2868381	1.607	0.861	28.574
	1.925	5.425	8.513	2.278	1.2868384	1.636	0.860	28.477
2	2.037	0.407	6.234	2.130	1.4108256	2.109	1.005	27.049
	1.928	0.073	6.887	2.095	1.4108281	2.014	1.011	27.300
3	2.614	5.399	7.001	1.997	1.0496413	0.657	4.379	25.834
	2.258	3.236	9.175	1.611	1.0496403	0.670	4.302	25.807
4	0.973	3.079	3.562	2.159	1.4499807	1.822	2.603	21.913
	0.653	3.163	3.134	1.562	1.4499779	1.919	2.549	22.433
5	0.794	1.146	7.543	2.447	0.9077304	1.899	3.754	18.679
	1.027	0.661	9.381	2.498	0.9077328	2.005	3.712	18.998
6	0.169	1.721	4.138	2.258	1.0125854	2.527	3.814	17.991
	1.229	2.567	9.616	1.386	1.0125906	2.546	3.472	18.413
7	0.707	5.536	7.505	0.378	0.8900178	2.181	2.625	17.933
	0.736	3.146	7.238	1.572	0.8900200	2.120	2.575	18.049
8	1.374	5.872	6.656	1.307	1.1128036	2.440	2.393	14.569
	1.469	3.157	8.624	2.892	1.1128001	2.445	2.528	15.000
9	2.265	5.416	4.265	2.420	1.0739092	1.861	2.984	14.004
	1.894	4.639	6.736	2.248	1.0739122	1.966	3.106	14.902
10	1.334	5.966	6.307	0.062	1.1657974	0.910	4.236	11.869
	1.200	5.440	4.746	0.123	1.1658040	0.681	4.120	12.613
11	1.292	6.011	8.820	2.480	0.9523456	1.860	5.740	12.057
	1.102	6.610	6.147	2.324	0.9523423	2.015	5.588	12.619
12	1.505	1.234	9.139	2.217	0.8539045	1.698	3.036	11.698
	1.433	4.555	9.850	0.661	0.8539042	1.762	3.109	11.809
13	1.310	1.508	2.749	0.964	1.4916823	1.821	3.124	11.179
	1.382	2.009	3.192	1.078	1.4916762	1.835	3.163	11.591
14	1.037	0.901	2.971	0.590	1.2312381	0.796	0.794	11.129
	1.076	3.150	3.480	1.598	1.2312356	0.785	0.722	11.516
15	1.633	2.941	7.499	1.073	1.0349592	2.173	4.157	10.819
	1.639	6.147	7.950	2.585	1.0349597	2.198	4.107	11.042
16	1.234	1.502	3.227	2.230	1.2654466	1.224	0.927	11.207
	1.290	1.607	4.010	2.000	1.2654471	1.264	0.955	11.755
17	1.885	0.241	6.627	1.720	0.8813608	2.248	6.083	10.498
	1.805	2.752	9.734	0.053	0.8813639	2.244	6.218	10.917
18	1.642	2.812	4.951	1.350	1.3962156	2.417	6.095	10.585
	1.389	3.152	3.461	0.663	1.3961841	1.188	1.141	10.881

TABLE II: True values (upper line) and recovered median values (lower line) for data set 1.

- Symposium, Astronomical Society of the Pacific Conference Series* (2013), vol. 467, p. 47.
- [6] N. J. Cornish and J. Crowder, Phys.Rev.D **72**, 043005 (2005).
- [7] P. Amaro-Seoane, S. Aoudia, S. Babak, P. Bintruy, E. Berti, A. Bohe, C. Caprini, M. Colpi, N. J. Cornish, K. Danzmann, et al., Classical Quantum Gravity **29**, 124016 (2012).
- [8] N. J. Cornish and S. L. Larson, Phys.Rev.D **67**, 103001 (2003).
- [9] A. Blaut, S. Babak, and A. Krolak, Phys.Rev.D **81**, 063008 (2010).
- [10] J. T. Whelan, R. Prix, and D. Khurana, Classical Quantum Gravity **27**, 055010 (2010).
- [11] N. J. Cornish and E. K. Porter, Classical Quantum Gravity **22**, S927 (2005).
- [12] J. Crowder, N. J. Cornish, and L. Reddinger, Phys.Rev.D **73**, 063011 (2006).
- [13] S. D. Mohanty and R. K. Nayak, Phys.Rev.D **73**, 083006 (2006).
- [14] M. Trias, A. Vecchio, and J. Veitch, Classical Quantum Gravity **26**, 204024 (2009).
- [15] J. Crowder and N. Cornish, Phys.Rev.D **75**, 043008 (2007).
- [16] N. J. Cornish and E. K. Porter, Classical Quantum Gravity **24**, 5729 (2007).
- [17] N. J. Cornish and E. K. Porter, Classical Quantum Gravity **24**, S501 (2007).
- [18] A. Vecchio and E. D. Wickham, Phys.Rev.D **70**, 082002 (2004).
- [19] A. Stroeer, J. Gair, and A. Vecchio, AIP Conf.Proc. **873**, 444 (2006).
- [20] J. Veitch and A. Vecchio, Phys.Rev.D **78**, 022001 (2008).
- [21] N. J. Cornish and T. B. Littenberg, Phys.Rev.D **76**, 083006 (2007).
- [22] T. B. Littenberg and N. J. Cornish, Phys.Rev.D **80**, 063007 (2009).
- [23] J. Kennedy and R. Eberhart, IEEE International Conference on Neural Networks Proceedings **4**, 1942 (1995).
- [24] Y. Shi and R. Eberhart, The 1998 IEEE International

Binary	i (rad)	ϕ_0 (rad)	A ($\times 10^{-22}$)	ψ (rad)	f_0 (mHz)	θ (rad)	ϕ (rad)	ρ
1	2.724	2.486	9.190	0.071	1.0137261	2.486	3.997	33.372
	2.497	3.161	10.258	1.565	1.0137248	2.467	3.987	33.398
2	0.657	5.360	9.485	0.157	0.9914607	1.007	4.245	31.033
	0.634	3.157	9.512	1.569	0.9914620	0.970	4.190	31.117
3	0.490	0.466	8.480	2.625	1.0149462	2.510	3.019	29.325
	0.759	3.130	9.926	1.599	1.0149459	2.490	3.027	29.347
4	0.654	5.595	8.554	0.232	1.0107870	1.179	4.783	29.059
	0.656	3.140	8.878	1.571	1.0107844	1.234	4.760	29.134
5	2.710	5.245	6.090	1.250	0.9907225	1.653	4.248	26.531
	2.499	3.159	7.889	1.568	0.9907217	1.688	4.258	26.581
6	0.087	3.574	6.578	1.917	0.9879189	1.517	1.154	25.948
	0.761	3.098	8.463	1.604	0.9879187	1.502	1.136	25.968
7	0.642	0.205	6.039	1.474	1.0129264	1.694	2.440	22.489
	0.685	3.125	6.448	1.560	1.0129254	1.804	2.430	22.519
8	2.435	1.804	5.927	1.392	0.9962224	1.614	5.307	21.139
	2.363	3.160	6.690	1.579	0.9962234	1.575	5.335	21.217
9	0.416	1.879	5.615	2.911	1.0132963	1.057	0.906	21.055
	0.692	3.127	6.224	1.600	1.0132978	1.079	0.948	20.871
10	0.573	0.381	5.691	1.367	0.9969600	1.705	5.894	18.579
	0.993	0.325	7.371	1.546	0.9969575	1.684	5.814	18.698
11	1.109	6.107	8.611	2.887	0.9932594	0.991	3.671	18.766
	1.128	6.293	9.225	2.788	0.9932599	0.988	3.661	18.778
12	2.918	3.411	5.474	1.230	0.9936598	2.416	3.223	18.298
	2.291	3.209	6.835	1.601	0.9936605	2.372	3.157	18.344
13	0.587	4.908	5.728	1.846	1.0010903	2.781	4.516	17.407
	0.866	3.097	6.929	1.525	1.0010910	2.839	4.481	17.517
14	2.909	5.736	4.362	1.583	1.0105097	1.915	4.873	17.207
	2.414	3.059	5.489	1.617	1.0105107	1.989	4.873	17.292
15	2.355	2.960	6.865	0.269	0.9997605	2.801	5.002	16.237
	2.090	3.164	8.168	1.543	0.9997629	2.842	5.003	16.388
16	0.763	5.151	5.300	3.043	1.0006721	1.475	4.290	15.903
	1.110	4.872	7.285	3.084	1.0006747	1.566	4.321	16.135
17	1.382	4.964	8.471	1.291	1.0087702	1.115	0.025	15.684
	1.334	4.871	9.959	1.292	1.0087716	1.062	0.017	15.791
18	2.413	4.630	4.752	0.491	0.9867635	1.230	4.405	15.853
	2.157	4.796	6.555	0.566	0.9867638	1.168	4.404	15.931
19	2.226	0.092	5.368	2.827	1.0063174	2.011	0.955	15.081
	2.384	3.134	4.746	1.570	1.0063165	2.058	0.949	15.104
20	0.714	1.599	5.244	0.996	0.9849761	1.785	0.078	15.473
	0.749	3.038	5.301	1.501	0.9849765	1.681	0.068	15.604
21	2.519	5.564	4.728	0.527	0.9966749	1.188	6.056	14.957
	2.179	3.109	5.907	1.568	0.9966725	1.215	6.029	14.838
22	1.400	3.590	9.501	2.058	1.0037884	1.387	1.419	15.368
	1.354	0.594	9.750	0.408	1.0037880	1.242	1.459	15.643
23	2.336	1.589	5.933	1.244	0.9944868	2.164	2.792	14.590
	2.314	3.114	5.099	1.569	0.9944899	2.131	2.710	14.644
24	1.068	0.755	6.286	0.107	0.9874341	1.749	2.011	14.236
	0.822	3.113	4.832	1.572	0.9874327	1.726	2.039	14.274
25	1.484	4.598	7.791	1.088	0.9987247	1.759	3.263	13.653
	1.505	1.332	9.130	2.646	0.9987257	1.660	3.227	13.795
26	0.598	0.256	3.732	0.489	1.0143321	1.709	4.371	13.398
	0.859	3.167	4.538	1.584	1.0143343	1.768	4.295	13.587
27	2.598	5.467	3.224	0.452	1.0092818	0.807	5.125	12.370
	2.172	3.135	5.185	1.577	1.0092834	0.678	5.351	12.774
28	1.354	4.636	7.251	1.196	0.9898903	1.722	0.786	12.877
	1.236	5.007	6.642	1.146	0.9898880	1.816	0.754	13.043
29	1.648	0.571	7.034	0.244	0.9983514	0.642	2.039	12.022
	1.578	3.455	7.034	1.850	0.9983527	0.631	2.197	12.055
30	2.713	4.060	3.253	0.771	1.0045681	2.501	5.855	10.966
	2.222	3.121	4.273	1.584	1.0045687	2.517	5.928	10.968

TABLE III: True values (upper line) and recovered median values (lower line) for data set 2.

- Conference on Evolutionary Computation Proceedings, IEEE Computer Society, Washington DC, USA. p. 69 (1998).
- [25] R. Storn and K. Price, *J. Global Opt.* **11**, 341 (1997).
 - [26] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, *J. Chem. Phys.* **21**, 1087 (1953).
 - [27] W. K. Hastings, *Biometrika* **57**, 97 (1970).
 - [28] M. G. H. Omran, A. P. Engelbrecht, and A. Salman, *Eur. J. Oper. Res.* **196**, 128 (2009).
 - [29] J. Prasad and T. Souradeep, *Phys.Rev.D* **85**, 123008 (2012), [Erratum: *Phys. Rev.D*90,no.10,109903(2014)].
 - [30] R. A. Hulse and J. H. Taylor, *Astrophys.J.* **195**, L51 (1975).
 - [31] J. Aasi et al. (LIGO Scientific Collaboration), *Classical Quantum Gravity* **32**, 115012 (2015), 1411.4547.
 - [32] F. Acernese et al. (VIRGO), *Classical Quantum Gravity* **32**, 024001 (2015).
 - [33] URL <http://www.ipta4gw.org/>.
 - [34] S. Nissanke, M. Vallisneri, G. Nelemans, and T. A. Prince, *Astrophys.J.* **758**, 131 (2012).
 - [35] E. K. Porter and N. J. Cornish, *Phys.Rev.D* **91**, 104001 (2015).
 - [36] C. Cutler, *Phys.Rev.D* **57**, 7089 (1998).
 - [37] N. J. Cornish and L. J. Rubbo, *Phys.Rev.D* **67**, 022001 (2003), [Erratum: *Phys. Rev.D*67,029905(2003)].
 - [38] A. Krolak, M. Tinto, and M. Vallisneri, *Phys.Rev.D* **70**, 022003 (2004), [Erratum: *Phys. Rev.D*76,069901(2007)].
 - [39] S. E. Timpano, L. J. Rubbo, and N. J. Cornish, *Phys.Rev.D* **73**, 122001 (2006).
 - [40] A. Petiteau, Y. Shang, S. Babak, and F. Feroz, *Phys.Rev.D* **81**, 104016 (2010).
 - [41] J. R. Gair and E. K. Porter, *Classical Quantum Gravity* **26**, 225004 (2009).
 - [42] C. ter Braak, *Stat. Comp.* **16**, 239 (2006).
 - [43] C. ter Braak and J. Vrugt, *Stat. Comp.* **18**, 435 (2008).
 - [44] S. R. Taylor, J. R. Gair, and L. Lentati, arXiv:1210.3489 (2012).
 - [45] Y. Wang, S. D. Mohanty, and F. A. Jenet, *Astrophys. J.* **795**, 96 (2014).
 - [46] Y. Wang and S. D. Mohanty, *Phys.Rev.D* **81**, 063002 (2010).
 - [47] URL http://www.particleswarm.info/Standard_PS0_2006.c.
 - [48] C. Huwyler, E. K. Porter, and P. Jetzer, *Phys.Rev.D* **91**, 024037 (2015).
 - [49] N. J. Cornish and E. K. Porter, *Phys.Rev.D* **75**, 021301 (2007).
 - [50] G. H. A. Roelofs, A. Rau, T. R. Marsh, D. Steeghs, P. J. Groot, and G. Nelemans, *Astrophys. J.* **711**, L138 (2010).
 - [51] E. K. Porter, *Phys.Rev.D* **92**, 064001 (2015).